



*Nikola Ebenbeck*



*Sven Anderson*



*Jana Jungjohann*



*Markus Gebhardt*

# Mehr Zeit für pädagogische Handlungen durch adaptives Testen

## Voraussetzungen, Chancen und Grenzen am Beispiel einer inklusiven Lesediagnostik

### Zusammenfassung

Digitale Instrumente ermöglichen eine einfache Testdurchführung und hohe Individualisierung im schulischen Alltag. Insbesondere adaptives Testen führt zu einer Zeiteffizienz beim Messen von Schülerinnen und Schülern. In diesem Beitrag beschreiben wir den Mehrwert des adaptiven Testens für Lehrkräfte und Schülerinnen und Schüler. Hierfür vergleichen wir die Durchführungszeit und Messgenauigkeit eines adaptiven und eines nicht-adaptiven Lesescreenings am Beispiel von 357 Drittklässlerinnen und -klässlern inklusiver Grundschulen. Adaptives Testen halbiert die Testdauer, wodurch mehr Zeit für pädagogische Handlungen im Unterricht zur Verfügung steht. Dies unterstützt die unterrichtsimmanente Diagnostik von Schülerinnen und Schülern mit sonderpädagogischem Unterstützungsbedarf. Die Einbettung adaptiver Tests in den Schulalltag durch digitale Lernumgebungen wird diskutiert.

Pädagogische Diagnostik begründet pädagogische Maßnahmen, die das Ziel haben, Personen zu helfen und zu unterstützen. Sie spielt eine entscheidende Rolle, um zu beurteilen, ob pädagogische Maßnahmen wirken oder ausreichen (Gebhardt, 2023). Bereits Klauer (1975) sieht die zentrale Aufgabe der pädagogischen Diagnostik darin, Planungs-, Handlungs-, Feststellungs- und Bewertungsentscheidungen zu legitimieren. Pädagogische Diagnostik ist dann inklusiv, wenn sie den Bedürfnissen und der Teilhabe aller Schülerinnen und Schüler sowie ihrer akademischen und sozialen Entwicklung dient und durch die Ergebnisse der Diagnostik wichtige pädagogische Entscheidungen im Alltag begründet und evaluiert werden können (Neumann & Lütje-Klose, 2020). Inklusiv Diagnostik muss entsprechend ein Prozess sein, der wiederkehrend im Schulalltag durchgeführt wird, um die Passung von individuellen Lernausgangslagen und pädagogischen Maßnahmen sicherzustellen. Damit inklusive Diagnostik zu wirksamen Maßnahmen führen kann, muss diese im Unterrichtsalltag möglichst einfach integrierbar sein und bestenfalls die Individualität der Schülerinnen und Schüler berücksichtigen (Jungjohann et al., 2022). Eine Möglichkeit hierfür ist der Einsatz von adaptiven Tests.

### Von der Feststellungsdiagnostik zur inklusiven Diagnostik

In der inklusiven Schule werden unterschiedliche diagnostische Tests eingesetzt, um verschiedene pädagogische Fragestellungen zu beantworten (Jungjohann & Gebhardt, 2023). Das Wort „Diagnostik“ wird in der Praxis selten mit pädagogischen Handlungen, sondern häufiger mit umfassenden Einzelfallgutachten verknüpft. Das liegt daran, dass die traditionelle sonderpädagogische Diagnostik auf Feststellungsdiagnostik (KMK, 2019, S. 11) und die damit verbundene Erstellung eines Gutachtens ausgelegt war (Bundschuh & Winkler, 2014). Einzelfallgutachten, wie sie zur Feststellung eines sonderpädagogischen Unterstützungsbedarfs angefertigt werden, dienen der Kategorisierung von Schülerinnen und Schülern. Sie sind in Deutschland rechtlich notwendig, um Entscheidungen, wie beispielsweise den Besuch einer Förderschule, zu begründen.

Hierfür werden Schulleistungs- oder Intelligenztests durchgeführt und ein sonderpädagogisches Einzelfallgutachten mit Hinweisen zu Umfang und Art der Unterstützung angefertigt (Gebhardt & Jungjohann, 2020).

Für die pädagogische Praxis sind diese Verfahren jedoch sehr langwierig und haben einen geringen Nutzen für die Förderplanung. So zeigen die Arbeiten von Beckstein und Sroka (2022) und Lutz (2022), dass viel Zeit für Feststellungsdiagnostik im Rahmen des Mobilen Sonderpädagogischen Diensts in Bayern aufgewendet wird und dadurch weniger zeitliche Ressourcen für Förderdiagnostik zur Verfügung stehen. Förderdiagnostik (KMK, 2019, S. 10) ist pädagogische Diagnostik, die zur Planung individueller Bildungsangebote und Fördermaßnahmen eingesetzt wird (Heimlich, 2012). Sie verzahnt Diagnostik und Förderung, um Schülerinnen und Schülern sonderpädagogische Förderung zu ermöglichen, die deren Lernfortschritt datengestützt im Blick hat.

In einem inklusiven Bildungssystem ist bestenfalls keine Kategorisierung von Schülerinnen und Schülern nötig und Ressourcen können stattdessen stärker für Förderdiagnostik genutzt werden. Die Durchführung von Tests dient dann dazu, eine sichere Datengrundlage zu bieten, mit deren Hilfe angemessene Förderentscheidungen für den jeweiligen Leistungsstand der Schülerinnen und Schüler getroffen und evaluiert werden können. Dabei werden sowohl die Lernausgangslage als auch die Lernentwicklung der Schülerinnen und Schüler in die Förderplanung einbezogen. Das erfordert die Entwicklung und den Einsatz neuer Verfahren und Testinstrumente, wie z. B. kurze Screenings für die Einschätzung des Leistungsstands und Prozessdiagnostik zur Analyse des Lernverlaufs (Gebhardt et al., 2022).

Screenings erheben den Lernstand der gesamten Klasse mittels einer kurzen Messung zu einem Messzeitpunkt (z. B. am Schuljahresanfang). Sie sind kürzer und enthalten weniger Subtests als umfangreiche Schulleistungstests zur Feststellungsdiagnostik (Buchwald et al., 2022a). Screenings werden genutzt, um Lernschwierigkeiten und Unterstützungsbedarf in einem Lernbereich zu identifizieren und Fördermaßnahmen zu legitimieren (Glover & Albers, 2007). Die Prozessdiagnostik nutzt informelle Verfahren oder Lernverlaufdiagnostik, um anhand mehrerer Messzeitpunkte einen Lernverlauf darzustellen. Sie beantwortet die Frage, ob der aktuelle Unterricht und die Förderung zum Erreichen der Förderziele führen. Sowohl Screenings als auch Instrumente für Prozessdiagnostik haben eine vergleichsweise kurze Durchführungszeit von teilweise nur wenigen Minuten. Sie benötigen im Unterrichtsalltag daher weniger zeitliche Ressourcen als Feststellungsdiagnostik (Buchwald et al., 2022b) und bieten zugleich mehr pädagogischen Nutzen. Die kurze Durchführungszeit wirkt sich insbesondere für Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf vorteilhaft aus. Sie profitieren vermutlich von diesen kurzen Tests, weil sie eher zum vorzeitigen Abbruch von Tests tendieren (List, 2018) sowie häufiger eine niedrige Frustrationstoleranz und geringe Motivation beim Erlernen schulischen Inhalte aufweisen (Lai, 2011).

In der Testentwicklung ist darauf zu achten, dass das Testinstrument einfach durchzuführen, zu analysieren und auszuwerten ist. Es sollte den anvisierten Kompetenzbereich bei allen Schülerinnen und Schülern einer inklusiven Lerngruppe zuverlässig und genau messen und beispielsweise Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf nicht benachteiligen. Im Vergleich dazu messen viele Testinstrumente im mittleren Kompetenzbereich genauer als in den Leistungsrandbereichen. Um das zu vermeiden, sollten Testinstrumente für den inklusiven Unterricht genügend leichte Aufgaben in den unteren Leistungsrandbereichen enthalten oder flexibel erweiterbar sein, um auch Schulleistungen von Schülerinnen und Schülern mit geringeren Kompetenzen reliabel messen zu können. Für den Einsatz im Unterrichtsalltag sollte das Testinstrument wenig Zeit beanspruchen, leicht durchführbar und interpretierbar sind und somit eine hohe Benutzerfreundlichkeit für Lehrkräfte haben. Das kann durch eine genaue Beschreibung der Regeln zur Durchführung und klare Instruktionen, Einsicht in die Konstruktionsprinzipien (z. B. mittels Schwierigkeitsmerkmalen, Anderson et al., 2022) sowie eine größere Anzahl an Beispielen erreicht werden.

## Digitalisierung als Chance für inklusive Diagnostik

## Adaptives Testen zur individuellen Messung von Schülerinnen und Schülern

Digitale Testinstrumente (sogenannte Computerbasierte Tests, CBT) greifen viele dieser Anforderungen auf und bieten Vorteile bei der Durchführung, Auswertung und Dokumentation von pädagogischer Diagnostik (Gebhardt et al., 2016). CBT werden auf einem digitalen Endgerät (z. B. Computer oder Tablet) durchgeführt. Sie steuern den Testdurchlauf und werten die Testergebnisse automatisch aus, um die Lehrkraft bei der Einhaltung der Regeln zur Durchführung und zur Auswertung zu entlasten. Automatische Testdurchführungen und -auswertungen führen zu einer höheren psychometrischen Qualität, da eine bessere Objektivität, Ökonomie und Benutzerfreundlichkeit für Lehrkräfte sowie Testfairness für Schülerinnen und Schüler erreicht werden können (Liebers et al., 2019). Fehler während der Durchführung, als auch Testleitereffekte, sprich Einflüsse, die die Lehrkraft unbewusst auf die Schülerinnen und Schüler während einer Testung ausüben kann und die deren Reaktionen und Ergebnisse beeinflussen, oder Fehler während der Durchführung können dadurch reduziert werden (Schaper, 2009; Walter & Schuhfried, 2004). Mit automatischen Auswertungen werden Übertragungs- oder Auswertungsfehler im Vergleich zu einer händischen Auswertung minimiert. Insgesamt werden so die Durchführung und Auswertung der Tests im schulischen Alltag beschleunigt und die Messungen nehmen weniger Zeit und Ressourcen in Anspruch.

Adaptive Tests sind CBT, die ihren Schwierigkeitsgrad automatisch an die Fähigkeiten der Schülerinnen und Schüler anpassen (van der Linden & Glas, 2000). Sie ermöglichen eine genauere und effizientere Messung von Schülerinnen und Schülern im Leistungsrandbereich, wie z. B. von Schülerinnen und Schülern mit sonderpädagogischem Unterstützungsbedarf (Ebenbeck, 2023). Bisher sind hauptsächlich nicht-adaptive („lineare“) Tests in Schulen verbreitet. In linearen Tests bearbeiten Schülerinnen und Schüler die Testaufgaben in einer meistens festgelegten Reihenfolge. Unterschiedliche Fähigkeitsniveaus der Schülerinnen und Schüler werden nicht berücksichtigt, weswegen alle dieselben Testaufgaben bearbeiten.

In einem adaptiven Test bearbeiten Schülerinnen und Schüler im Vergleich dazu nicht dieselben Items. Stattdessen bekommt jede Person automatisch eine individuelle Itemauswahl auf ihrem Leistungsniveau. Die automatisierte Itemauswahl wird durch einen vorab programmierten Algorithmus gesteuert, der für jede Person passende Aufgaben basierend auf ihren bereits beantworteten Aufgaben auswählt. Zu leichte oder zu schwierige Aufgaben werden aussortiert und müssen von der Person nicht beantwortet werden. Dadurch verkleinert sich die Gesamtanzahl von Testaufgaben, die eine Person bearbeiten muss und die Testzeit wird verkürzt. Die Testgenauigkeit und Reliabilität bleiben jedoch ähnlich wie bei linearen Tests (Stone & Davey, 2011). Für Schülerinnen und Schüler mit schwachen Leistungen bedeutet das, dass sie tendenziell mehr Testaufgaben bekommen, die sie potenziell lösen können, anstatt überwiegend Testaufgaben zu erhalten, die für sie viel zu schwierig sind. Adaptives Testen führt damit zu weniger Über- bzw. Unterforderung dieser Schülerschaft und theoretisch zu einer höheren Testmotivation (Akhtar et al., 2023, Martin & Lazendic, 2018).

Es gibt eine kleine Auswahl an adaptiven Tests, die in Deutschland aktuell eingesetzt werden kann, wie z. B. der adaptive Intelligenztest AID-3 bzw. AID-3 tailored (Kubinger & Holoher-Benetka, 2023) oder der digitale adaptive Test FIPS für die erste Jahrgangsstufe (Bäuerlein et al., 2021). Im amerikanischen Raum wird adaptives Testen vor allem in universitären und beruflichen Auswahlverfahren eingesetzt (z. B. Rudner, 2007, Weiss & Kingsbury, 1984) oder als schulische Lernverlaufdiagnostik, beispielsweise im adaptiven Testsystem Star Math verwendet (Shapiro et al., 2015). Die überschaubare Auswahl an adaptiven Testinstrumenten liegt insbesondere an deren erheblichen Aufwand der Testentwicklung und -programmierung. Für die Entwicklung adaptiver Tests ist eine große Anzahl an Testaufgaben notwendig, deren Schwierigkeit sich im besten Fall über alle Fähigkeitsniveaus erstreckt (Way, 2005, Weiss & Kingsbury, 1984). Um einen neu entwickelten adaptiven Test prüfen und skalieren zu können, müssen mindestens 250 Schülerinnen und Schülern alle Testaufgaben bearbeiten (Linacre, 1994). Dabei sollten auch Schülerinnen und Schülern mit sonderpädagogischem Unterstützungsbedarf berücksichtigt werden, um die Fairness des adaptiven Tests sicherzustellen. Alternativ können bereits testtheoretisch geprüfte

lineare Testinstrumente aus dem sonderpädagogischen Bereich um einen adaptiven Algorithmus erweitert werden. Ebenso können mehrere verschiedene Testinstrumente mit wenigen Items zu einem größeren Testinstrument mit vielen Items auf verschiedenen Schwierigkeitsstufen kombiniert werden (Ebenbeck, 2023; Ebenbeck & Gebhardt, 2023).

Neben der Erstellung eines geeigneten Itempools sind weitere Schritte notwendig, um einen adaptiven Test zu entwickeln. Diese Schritte werden nachfolgend beispielhaft an der Entwicklung des adaptiven Lesescreenings LES-IN-CAT (LEseScreening - INklusiver Computerbasierter Adaptiver Test) für die Testplattform Levumi.de dargestellt. Levumi.de ist eine Webseite, die diagnostische Tests und Fördermaterialien als Open Educational Resources (OER) der Schullandschaft kostenfrei anbietet (Jungjohann et al., 2018). Die Entwicklung des adaptiven Tests fand im vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projekt DaF-L statt.

Im ersten Schritt wurde das analoge Lesescreening LES-IN (LEseScreening – INklusiv; Ebenbeck et al., 2022a), bestehend aus vier Subtests entwickelt, die jeweils einen Teilbereich der Lesefähigkeit (Phonologische Bewusstheit, Geschwindigkeit im lexikalischen Abruf, Sicherheit im lexikalischen Abruf und sinnkonstruierendes Satzlesen) im Klassenverband messen. Die Testaufgaben der Subtests wurden in einer querschnittlichen und bundeslandübergreifenden Schulstudie mit 709 Schülerinnen und Schülern in Grund- und Förderschulen psychometrisch geprüft (Jungjohann et al., im Druck). Im zweiten Schritt wurde basierend auf dem analogen Lesescreening die digitale Weiterentwicklung LES-IN-DIG (LEseScreening - INklusiv und DIGital, Ebenbeck et al., 2022b) auf Levumi.de implementiert und in einer weiteren Schulstudie mit 400 Grund- und Förderschülerinnen und -schülern geprüft (Ebenbeck, 2023). Im dritten Schritt wurden verschiedene adaptive Algorithmen programmiert und empirisch verglichen, um die passendste Möglichkeit für die Zielgruppe und den betreffenden Test zu finden (Ebenbeck, 2023). Dabei wurde auf Testergebnisse der Stichprobe aus Schritt zwei zurückgegriffen, um adaptive Tests vor der aufwendigen Programmierung des eigentlichen adaptiven Tests zu simulieren. Mit Simulationsstudien kann z. B. überprüft werden, wie genau der adaptive Test gewisse Schülergruppen messen würde, welche Testaufgaben konkret für welche Schülerinnen und Schüler gezogen werden und wie lange ein Testdurchlauf in etwa dauern würde.

Die ausführlichen Analysen zum Test LES-IN-CAT sind in der Dissertation von Ebenbeck (2023) dargestellt. Für diesen Artikel werden die wichtigsten Ergebnisse anhand einer kleineren Stichprobe mit zusätzlichen weiterführenden Analysen zur Auswahl der Testaufgaben anhand des Subtests zur phonologischen Bewusstheit vorgestellt. Leitende Fragestellungen sind dabei:

- Inwieweit verkürzt ein adaptiver Test die Messung von Schülerinnen und Schülern mit und ohne sonderpädagogischen Unterstützungsbedarf?
- Welche Testaufgaben werden für Schülerinnen und Schüler mit geringen Leistungen ausgewählt bzw. nicht ausgewählt?

Zwischen November 2022 und Januar 2023 wurden 357 Schülerinnen und Schüler der Jahrgangsstufen 2 bis 4 an inklusiven Grundschulen in Nordrhein-Westfalen mit dem digitalen linearen Lesescreening LES-IN-DIG (Ebenbeck et al., 2023) gemessen. Die Schülerinnen und Schüler waren durchschnittlich 8,43 Jahre alt ( $SD=1,07$ ). 43 Schülerinnen und Schüler hatten einen sonderpädagogischen Unterstützungsbedarf im Bereich Lernen oder Sprache. Sie bearbeiteten das Screening webbasiert auf Tablets im Klassenverbund. Als Beispiel wird im Folgenden ausschließlich der erste von vier Subtests ausgewertet, welcher in der linearen Version 35 Items umfasst und auf eine maximale Bearbeitungszeit von fünf Minuten begrenzt ist. Als Testaufgaben sehen die Schülerinnen und Schüler ein Bild bzw. ein graphisch repräsentiertes Wort sowie einen schriftsprachlichen Laut. Sie entscheiden, ob der Laut am Anfang, in der Mitte, am Ende des Worts oder gar nicht im Wort zu hören ist (Abb. 1).

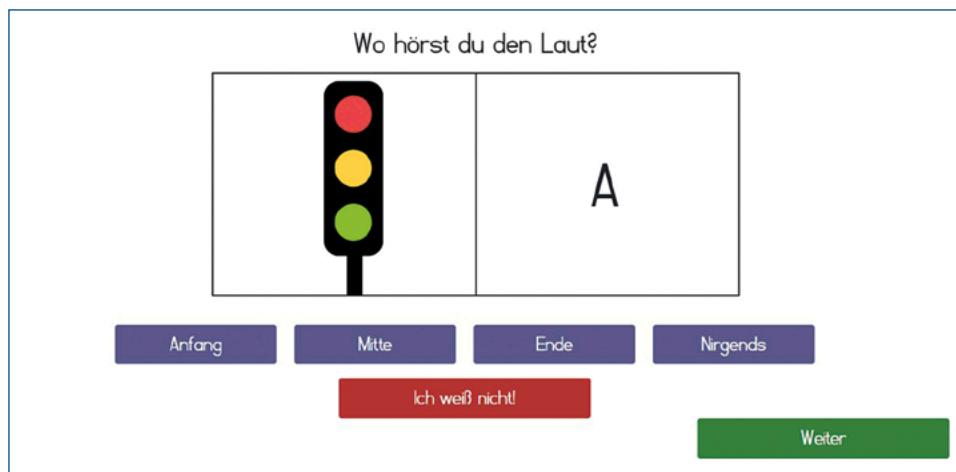
Um einen adaptiven Subtest basierend auf den Ergebnissen des linearen Subtests zu simulieren, wurden mehrere Schritte durchgeführt: Im ersten Schritt wurde der adaptive Algorithmus selbst

## Entwicklung eines adaptiven Lesescreenings

## Fragestellungen

## Methode

Abb. 1:  
Beispielaufgabe des Tests  
„Phonologische Bewusstheit“



aufgesetzt. Hierfür wurde die Programmiersprache R mit dem Paket catR (Magis & Barrada, 2017) verwendet, welches speziell der Simulation von adaptiven Tests dient. Es wurden zwei verschiedene Algorithmen für adaptives Testen verglichen. Im ersten Fall beginnt der adaptive Test mit einem Item mittlerer Schwierigkeit, welches im Rahmen der Item-Response-Theorie (IRT) das aussagekräftigste Item ist. Im zweiten Fall beginnt der adaptive Test mit einer Testaufgabe mit leichter Schwierigkeit, um für Schülerinnen und Schüler mit schwächeren Leistungen einen motivierenden Teststart zu ermöglichen und den Einstieg in die Messung zu erleichtern. In beiden Algorithmen wird festgehalten, dass nach jeder Antwort die Fähigkeit der Person basierend auf ihren bislang gelösten Aufgaben geschätzt wird. So kann das nächste Item individuell passend ausgewählt werden. In beiden Fällen endet der adaptive Test, sobald die Fähigkeit der Person mit einem Standardfehler von 0.5 genau geschätzt werden kann. Über einen geringeren Standardfehler würde sich die Genauigkeit der Messung erhöhen lassen. Das würde jedoch auch zu einer längeren Testung führen.

Im zweiten Schritt wurden die empirisch erhobenen Daten der Schülerinnen und Schüler genutzt, um deren Fähigkeiten und die Schwierigkeit der Testaufgaben zu bestimmen. Um Deckeneffekte der Messung für die Simulationen auszugleichen und dadurch ein möglichst realistisches Bild zu geben, wurden basierend auf der Fähigkeitsverteilung der Schülerinnen und Schüler neue Fähigkeitswerte für Schülerinnen mit und ohne sonderpädagogischen Unterstützungsbedarf generiert (n=1000).

Mit diesen Bausteinen (Algorithmus, Personenfähigkeiten und Aufgabenschwierigkeiten) wurde der Test zur phonologischen Bewusstheit als adaptiver Test mit zwei verschiedenen Einstellungen simuliert und mit dessen linearer Version verglichen. Es wurden die durchschnittliche Testlänge und die erreichte Testgenauigkeit jeweils für Schülerinnen und Schüler mit und ohne sonderpädagogischen Unterstützungsbedarf gegenübergestellt.

## Ergebnisse

### Verkürzung der Testzeit

Der simulierte adaptive Test ist im Durchschnitt um ca. 50% kürzer als die lineare Version des Tests (Abb. 2), was bei einer Testzeit von fünf Minuten eine Reduktion auf zwei bis drei Minuten bedeuten würde. Die Testverkürzung unterscheidet sich zwischen den angenommenen Schülergruppen. Für Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf ist die Reduktion mit 55% größer als für Schülerinnen und Schüler ohne Unterstützungsbedarf mit 48%. Der adaptive Test wäre demnach für Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf im Schnitt kürzer als für Schülerinnen und Schüler ohne Unterstützungsbedarf.

Trotz einer Verkürzung der Testlänge ändert sich die Genauigkeit des adaptiven Tests im Vergleich zur linearen Version kaum. Die Genauigkeit kann anhand der Korrelation zwischen der realen gemessenen und der durch den adaptiven Test geschätzten Fähigkeit der Person bemessen

werden. Der lineare Test misst Schülerinnen und Schüler ohne sonderpädagogischen Unterstützungsbedarf mit  $r = .91$  und Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf mit  $r = .94$ . Beim adaptiven Test ist diese Genauigkeit marginal geringer und beträgt für Schülerinnen und Schüler mit und ohne Unterstützungsbedarf ( $r = .90$ ). Ob der adaptive Test mit einem einfachen oder einer mittelschwierigen Aufgabe beginnt, hat keinen Einfluss auf die durchschnittliche Testlänge und Testgenauigkeit für Schülerinnen und Schüler mit und ohne sonderpädagogischem Unterstützungsbedarf.

### Aufgabenauswahl

Leistungsschwache Schülerinnen und Schüler würden im adaptiven Test ca. 35% aller Testaufgaben bearbeiten. Die durch den Algorithmus gewählten Aufgaben liegen, wie erwartet, im Bereich ihrer Fähigkeiten (Abb. 3). Für leistungsschwache Schülerinnen und Schüler mit einer Fähigkeit von -1 bedeutet das beispielsweise, dass nur Aufgaben mit einer Schwierigkeit zwischen -0,5 und -1,5 gezogen werden, wobei eine Schwierigkeit von 0 einer Aufgabe mittlerer Schwierigkeit entspricht und negative Werte auf einfachere Aufgaben hindeuten. Items mit höheren Schwierigkeitswerten würden für solche Schülerinnen und Schüler stattdessen gar nicht vorgeschlagen werden, da sie zu schwierig sind und daher nicht genügend Aussagekraft über deren Fähigkeit haben. Insgesamt kann der Test somit um 21 Items gekürzt werden, wobei vor allem schwierige Items entfallen.

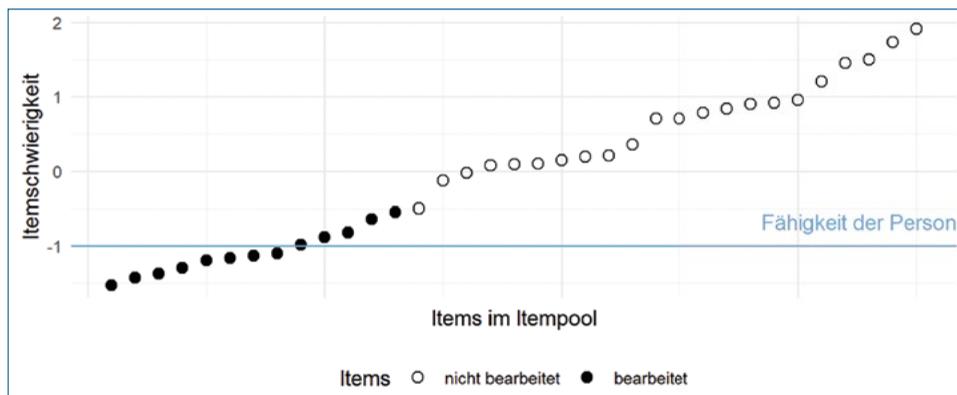


Abb. 3:  
Items, die eine Person mit der Fähigkeit -1 (eine Standardabweichung entfernt vom Durchschnitt) in einem beispielhaften simulierten adaptiven Testdurchlauf bearbeiten und nicht bearbeiten würde.

Adaptives Testen hat das Ziel, einen breiten Leistungsbereich in kürzerer Zeit zu messen als lineare traditionelle Tests, bei denen alle Items des Tests von den Schülerinnen und Schülern bearbeitet werden. In der vorliegenden Studie konnte dies am Subtest zur phonologischen Bewusstheit des Lesescreenings LES-IN-DIG (Ebenbeck et al., 2023) gezeigt werden. Dabei verdeutlichen die Ergebnisse, dass adaptives Testen die Testzeit insbesondere für Schülerinnen und Schüler mit sonderpädagogischem Unterstützungsbedarf reduziert, was sich mit den Überlegungen von Stone und Davey (2011) deckt.

Anstelle einer fünfminütigen Testdurchführung werden durch den Einsatz des adaptiven Testens lediglich zwei bis drei Minuten Testzeit benötigt. Diese Kürzung entlastet die Schülerinnen und Schüler bereits bei einem einzelnen Subtest. Bei umfangreichen Statustests, wie z. B. Schulleistungstests oder Intelligenztests – mit meist bis zu zwölf Dimensionen (Subtests) – hat eine Verkürzung der Testzeit für jeden einzelnen Subtest einen deutlichen Einfluss auf die Gesamtlänge des Tests und die Entlastung summiert sich auf. Wenn adaptives Testen in allen Subtests eines Statustests eingesetzt wird, könnte die Durchführungszeit des gesamten Statustests um bis zu 50% reduziert werden. Eine Testzeitverkürzung von 50% kann bei langen Statustests durchaus 30 bis 60 Minuten weniger Bearbeitungsdauer für Schülerinnen und Schüler bedeuten. Die Dauer eines Testdurchgangs wäre somit deutlich kürzer. Schülerinnen und Schüler werden weniger belastet und Lehrkräfte müssen weniger Unterrichts- und Förderzeit für Diagnostik aufwenden.

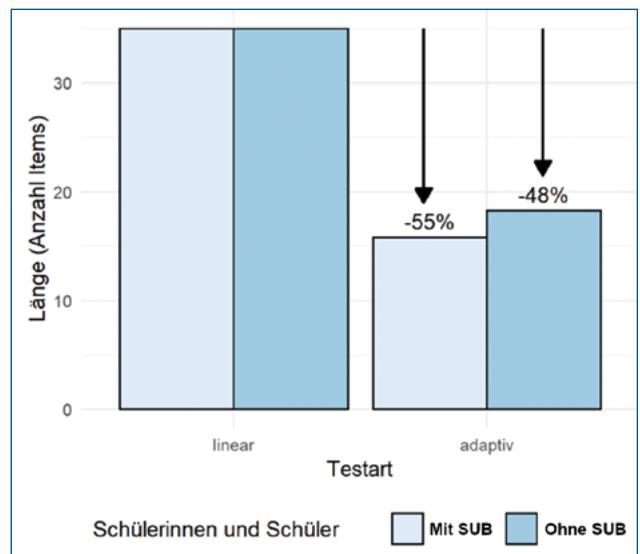


Abb. 2:  
Durchschnittliche Testlänge des ursprünglichen linearen Tests und des adaptiven Tests für Schülerinnen und Schüler mit und ohne sonderpädagogischen Unterstützungsbedarf (SUB) sowie erreichter prozentualer mittlerer Reduktion der Testlänge.

## Diskussion

Aufgrund der reduzierten Testlänge sinkt die Reliabilität und Genauigkeit des adaptiven Tests leicht. Dennoch eröffnet die verkürzte Testzeit im Rahmen des adaptiven Testens insbesondere für Schülerinnen und Schüler mit Lernschwierigkeiten oder sonderpädagogischem Unterstützungsbedarf neue Möglichkeiten, da diese besonders von kurzen Tests profitieren können (Lai, 2011, List, 2018). Diesen Schülerinnen und Schülern fällt es vermeintlich leichter, einen kürzeren adaptiven Test zu absolvieren, wodurch eine Verringerung der Abbruchquote von Testdurchläufen erzielt werden kann. Offen bleibt derzeit noch, inwieweit adaptive Tests sich auch positiv auf die Testmotivation auswirken (Akhtar et al., 2023, Martin & Lazendic, 2018).

Computerbasierte adaptive Screenings sind in der Lage, sowohl im oberen als auch im unteren Messbereich zuverlässige Ergebnisse zu liefern. Damit eignen sie sich besonders für die Messung von Kompetenzen in leistungsheterogenen Klassen. Insgesamt stellt die adaptive Ziehung eine Erweiterung für psychometrische Instrumente dar, da sie auf einem traditionellen linearen Test aufbaut. Einfach umzusetzen ist die Erweiterung von Tests durch Adaptivität, wenn als Grundlage ein digitaler Test verwendet wird. Die erfolgreiche Erstellung von adaptiven Tests erfordert jedoch einen umfangreichen Itempool und einen breiten Messbereich als Grundvoraussetzung (Way, 2005, Weiss & Kingsbury, 1984). Wenn Tests ungenau messen oder über einen zu engen Messbereich verfügen, kann eine adaptive Ziehung kontraproduktiv sein, da dann die Messung noch ungenauer wird. Daher ist es entscheidend, bei der Entwicklung adaptiver Tests auf einen breiten Anforderungsbereich der Testinhalte zu achten, um aussagekräftige Ergebnisse zu erzielen.

Schulleistungstests werden normalerweise einmal entwickelt, psychometrisch geprüft und dann in einer bestimmten Version vertrieben. Eine erneute Überprüfung oder Weiterentwicklung erfolgt meist erst nach vielen Jahren oder aus Lizenzgründen gar nicht. Werden digitale Lern- und Testplattformen für adaptives Testen genutzt, können weitere Schritte in der Testentwicklung bzw. -weiterentwicklung durchgeführt werden. So erfolgt zunächst ebenfalls der traditionelle Testprozess, der die Testkonstruktion, die Durchführung der psychometrischen Messung und die Skalierung als linearen Test umfasst. Danach wird der Test um einen adaptiven Algorithmus erweitert und als adaptiver Test in die Testplattform integriert. Ein adaptiver Test passt sich den individuellen Fähigkeiten der Schülerinnen und Schüler an, indem er gezielt Fragen auswählt, die

dem aktuellen Leistungsniveau entsprechen, wie unsere Ergebnisse zeigen konnten. Dadurch kann der Test eine effizientere Messung individueller Fähigkeiten ermöglichen. Im nächsten Schritt werden unter Berücksichtigung der adaptiven Testergebnisse Förderentscheidungen getroffen und die bisherige Skalierung des Tests weiterentwickelt. Die Antworten, die die Schülerinnen und Schüler während einer Förderung geben, können gemeinsam mit den Testergebnissen genutzt werden, um mit Hilfe adaptiver Algorithmen und automatischer Itemgenerierung maßgeschneiderte Unterstützung und individuelle Lernpläne für die Schülerinnen und Schüler zu erstellen. Durch diese kontinuierliche Datenerhebung und -analyse während Test und Förderung können Schwächen identifiziert und gezielte Fördermaßnahmen entwickelt werden, um das Lernen der Schülerinnen und Schüler zu verbessern. Die Testplattform ermöglicht somit einen iterativen Entwicklungsprozess, mit dem es gelingt, den Test kontinuierlich zu verbessern und an die Bedürfnisse der Schülerinnen und Schüler anzupassen. Durch diese fortlaufende Entwicklung stellt die Testplattform sicher, dass sie stets aktuell ist und den bestmöglichen Nutzen für Lehrkräfte und Schülerinnen und Schüler bietet.

Digitale bzw. adaptive Tests können mit Übungen zu Förderumgebungen kombiniert werden. Solche Test- und Förderumgebungen bieten eine verbesserte Einführung in die Handhabung der Testinstrumente im Vergleich zu traditionellen Tests. Dies wird durch erklärende Workshops, Videos und

## Schlüsselwörter

Pädagogische Diagnostik, Digitalisierung, Individualisierung, Adaptives Testen, Inklusion

## Abstract

Digital instruments allow for easy testing and high individualization in schools. In particular, adaptive testing leads to time-efficient measurement of students. In this paper we describe the added value of adaptive testing for teachers and students. We compare the administration time and measurement accuracy of an adaptive and a non-adaptive reading screening on a sample of 357 third graders from inclusive schools. Adaptive testing cuts testing time in half, leaving more time for instructional interventions in the classroom. This supports instructional diagnosis for students with disabilities. Embedding adaptive testing into the school day through digital learning environments is discussed.

## Keywords

Educational Assessment, Digitalisation, Individualisation, Adaptive Testing, Inclusion

Lehrhandbücher ermöglicht, die den Nutzerinnen und Nutzern eine klare Orientierung und Anleitung bieten (Jungjohann & Lutz, 2021). Durch diese zusätzlichen Hilfsmittel können die Tests und Fördermaßnahmen besser miteinander verknüpft und somit effektiver genutzt werden. Entsprechende Lern- und Testumgebungen sind auch bei der Entwicklung von der Statusdiagnostik hin zur Prozessdiagnostik erforderlich, da diese über die einmalige Statusdiagnostik und Gutachtererstellung hinausgehen, indem sie einen kontinuierlichen und dynamischen Diagnoseprozess ermöglichen. Neben den Tests umfassen diese Umgebungen auch Fördermaßnahmen, die gezielt auf die individuellen Bedürfnisse der Lernenden eingehen. Insgesamt ermöglichen digitale Test- und Förderumgebungen auf diese Weise so eine effiziente und zielgerichtete Diagnostik und Förderung.

Inklusive Schulen stehen vor der Herausforderung, pädagogische Diagnostik mit klaren Standards zu etablieren, um effektive Tests und (digitale) Lernumgebungen zu gestalten. Adaptive Screenings erfordern definierte Itempools und festgelegte Kompetenzprofile, die als Grundlage für die Entwicklung und Ausrichtung der Tests dienen sollen. Die technische Umsetzung bietet Möglichkeiten für digitale Diagnostik sowie Fördermaßnahmen und Lernspiele. Gegenwärtig mangelt es zum einen an den erforderlichen Standards und gewünschten Lernprofilen in Bildungseinrichtungen, um derartige Ansätze zu entwickeln. Zum anderen fehlen Anreize seitens staatlicher Stellen für die Wissenschaft und Unternehmen, um solche Systeme zu initiieren und zu pflegen. Die Nutzung digitaler Potenziale stellt somit eine der aktuellen Herausforderungen im Bildungssystem dar.

Akhtar, H., Silfiasari, Vekety, B., & Kovacs, K. (2023). The Effect of Computerized Adaptive Testing on Motivation and Anxiety: A Systematic Review and Meta-Analysis. *Assessment, 30*(5), 1379–1390.

Anderson, S., Sommerhoff, D., Schurig, M. & Gebhardt, M. (2022). Schwierigkeitsmerkmale als Grundlage für eine systematische und praxisnahe Konstruktion der Lernverlaufsdagnostik bei Additions- und Subtraktionsaufgaben im Hunderterraum. In S. Fränkel, M. Grünke, T. Henne- mann, D. C. Hövel, C. Melzer & K. Ziemer (Hrsg.), *Teilhabe in allen Lebensbereichen? Ein Blick zurück und nach vorn* (S. 81-86). Bad Heilbrunn: Julius Klinkhardt.

Bäuerlein, K., Beinicke, A., Schorr, M., Schneider, W. (2021). *Fähigkeitsindikatoren Primar- schule. Ein digitales Testverfahren zur Erfassung der Lernausgangslage und der Lernent- wicklung in der 1. Klasse. 2., aktualisierte und neu normierte Auflage*. Hogrefe.

Beckstein, G. & Sroka, S. (2022). Ein Erfahrungsbericht aus Bayern – Mobiler Sonderpädagogi- scher Dienst Sprache, Lernen, emotionale und soziale Entwicklung. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik*.

Buchwald, K., Ebenbeck, N. & Gebhardt, M. (2022a). Screenings, Status- und adaptive Tests in der schulischen Diagnostik. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozess- diagnostik und Förderplanung* (S. 305-312). Regensburg: Universitätsbibliothek. <https://doi.org/10.5283/epub.53149>

Buchwald, K., Anderson, S., Lutz, S., Mühling, A., Sommerhoff, D. & Gebhardt, M. (2022b). Lern- verlaufsdagnostik in Mathematik. Basiskompetenzen mit der Plattform Levumi.de messen. *Zeit- schrift für Heilpädagogik, 73*(4), 168–178.

Bundschuh, K. & Winkler, C. (2014). *Einführung in die sonderpädagogische Diagnostik* (UTB Sonderpädagogik, pädagogische Psychologie, Bd. 999, 8., überarbeitete Auflage). München: Ernst Reinhardt Verlag.

Ebenbeck, N. (2023). *Computerized Adaptive Testing in Inclusive Education* [Dissertation., Uni- versität Regensburg]. Publikationsserver der Universität Regensburg. [https://epub.uni-regensburg.de/54551/1/Ebenbeck%20-%20CAT%20in%20inclusive%20education\\_Final\\_26.07.2023.pdf](https://epub.uni-regensburg.de/54551/1/Ebenbeck%20-%20CAT%20in%20inclusive%20education_Final_26.07.2023.pdf)

Gebhardt, M. (2023). *Pädagogische Diagnostik. Leistung, Kompetenz und Entwicklung mes- sen, bewerten und interpretieren für individuelle Förderung*. (Version 0.2). Universität Regensburg. <https://doi.org/10.5283/epub.54450>

## Ausblick

## Literatur

- Gebhardt, M., Diehl, K. & Mühling, A. (2016). Online-Lernverlaufsmessung für alle SchülerInnen in inklusiven Klassen. www.LEVUMI.de. *Zeitschrift für Heilpädagogik*, 67(10), 444–454.
- Gebhardt, M., Scheer, D. & Schurig, M., (2022) *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung*. Version 1.0. Regensburger Beiträge zur Inklusions- und Sonderpädagogik, 4. Universität Regensburg. <https://doi.org/10.5283/epub.53149>
- Glover, T. A. & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, 45(2), 117–135. <https://doi.org/10.1016/j.jsp.2006.05.005>
- Jungjohann, J., DeVries, J. M., Gebhardt, M. & Mühling, A. (2018). Levumi: A web-based curriculum-based measurement to monitor learning progress in inclusive classrooms. In K. Miesenberger & G. Kouroupetroglou (Hrsg.), *Computers helping people with special needs. ICCHP 2018. Lecture notes in computer science* (S. 369-378). Springer International Publishing. [https://doi.org/10.1007/978-3-319-94277-3\\_58](https://doi.org/10.1007/978-3-319-94277-3_58)
- Jungjohann, J., & Lutz, S. (2021). *Schulische Förderpraxis mit und durch Lernverlaufsdagnostik. Spuren*, (2/2021), 40–45. <https://doi.org/10.5283/epub.45556>
- Jungjohann, J. & Gebhardt, M. (2023). Dimensions of Classroom-Based Assessments in Inclusive Education: A Teachers' Questionnaire for Instructional Decision-Making, Educational Assessments, Identification of Special Educational Needs, and Progress Monitoring. *International Journal of Special Education (IJSE)*, 38(1), 131–144. <https://doi.org/10.52291/ijse.2023.38.12>
- Jungjohann, J., Bastian, M., Mühling, A. & Gebhardt, M. (2022). Die Sicht von Lehrkräften auf die Implementation und den Nutzen von webbasierten Lernverlaufstests.: Eine Interviewstudie in inklusiven Grundschulen. In N. Harsch, M. Jungwirth, M. Stein, Y. Noltensmeier & N. Willenberg (Hrsg.), *Diversität Digital Denken – The Wider View.: Eine Tagung des Zentrums für Lehrerbildung der Westfälischen Wilhelms-Universität Münster vom 08. bis 10.09.2021* (S. 405-408). WTM-Verlag. <https://doi.org/10.37626/GA9783959871785.0.42>
- Heimlich, U. (2012). *Inklusive Bildung für Menschen mit Behinderung – eine lebenslaufbegleitende Perspektive*. <https://www.bildungsbericht.de/de/schwerpunktthemen/pdfs/expertenworkshop-2012-heimlich1112.pdf?msclid=d521b061c7a811ec90a719afa16deb99> [08.01.2024]
- Klauer, K. J. (1975). *Lernbehindertenpädagogik*. Halle: Carl Marhold.
- KMK, (2019). *Empfehlungen zur schulischen Bildung, Beratung und Unterstützung von Kindern und Jugendlichen im sonderpädagogischen Schwerpunkt Lernen*. [https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen\\_beschluesse/2019/2019\\_03\\_14-FS-Lernen.pdf](https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2019/2019_03_14-FS-Lernen.pdf)
- Kubinger & Holocher-Benetka (2023). *Adaptives Intelligenz Diagnostikum 3. 2.*, aktualisierte Auflage mit überprüften Normen. Göttingen: Hogrefe.
- Lai, E. R. (2011). Motivation: A literature review. *Person Research's Report*, 6, 40–41.
- Linacre, L. M. (1994). Sample Size and Item Calibration Stability. *Rasch Measurement Transactions* 7(4), 328
- List, M. K. (2018). *Testbearbeitungsverhalten in Leistungstests: Modellierung von Testabbruch und Leistungsabfall* [Doctoral dissertation, Christian-Albrechts Universität Kiel].
- Liebers, K., Kanold, E., & Junger, R. (2019). Digitale Lernstandsanalysen in der inklusiven Grundschule? In S. Bartusch, C. Klektau, T. Simon, & Teumer, S. Weidemann, A. (Hrsg.), *Lernprozesse Begleiten* (S. 209-221). Springer Fachmedien Wiesbaden. [https://doi.org/10.1007/978-3-658-21924-6\\_16](https://doi.org/10.1007/978-3-658-21924-6_16)
- Lutz, S. (2022). Mobile Sonderpädagogische Dienste (MSD) in Bayern. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 765-772). Regensburg: Universitätsbibliothek. <https://doi.org/10.5283/epub.53149>
- Magis, D., & Barrada, J. R. (2017). Computerized adaptive testing with R: Recent updates of the package catR. *Journal of Statistical Software*, 76, 1–19.
- Martin, A. J., & Lazendic, G. (2018). Computer-adaptive testing: Implications for students' achievement, motivation, engagement, and subjective test experience. *Journal of educational psychology*, 110(1), 27. <https://doi.org/10.1037/edu0000205>

- Neumann, P., & Lütje-Klose, B. (2020). Diagnostik in inklusiven Schulen - zwischen Stigmatisierung, Etikettierungs-Ressourcen-Dilemma und förderorientierter Handlungsplanung. In C. Gresch, P. Kuhl, M. Grosche, C. Sälzer & P. Stanat (Hrsg.), *Schüler\*innen mit sonderpädagogischem Förderbedarf in Schulleistungserhebungen. Einblicke und Entwicklungen* (S. 3-28). Wiesbaden: Springer VS.
- Rudner, L. M. (2007). Implementing the Graduate Management Admission Test® computerized adaptive test. In D. J. Weiss (Ed.), *Proceedings of the 2007 GMAC Conference on Computerized Adaptive Testing*. [www.psych.umn.edu/psylabs/CATCentral/](http://www.psych.umn.edu/psylabs/CATCentral/)
- Schaper, N. (2009). Online-Tests aus diagnostisch-methodischer Sicht. In H. Steiner (Hrsg.), *Online-Assessment* (S. 17-36). Berlin/Heidelberg: Springer. [https://doi.org/10.1007/978-3-540-78919-2\\_2](https://doi.org/10.1007/978-3-540-78919-2_2)
- Shapiro, E. S., Dennis, M. S., & Fu, Q. (2015). Comparing computer adaptive and curriculum-based measures of math in progress monitoring. *School Psychology Quarterly*, 30(4), 470.
- Stone, E., & Davey, T. (2011). Computer-Adaptive Testing for Students with Disabilities: A Review of the Literature. *ETS Research Report Series*, 2011(2), i-24. <https://doi.org/10.1002/j.2333-8504.2011.tb02268.x>
- Glas, C. A. W., & van der Linden, W. J. (Eds.). (Year). *Computerized Adaptive Testing: Theory and Practice*, 11-25. [https://doi.org/10.1007/0-306-47531-6\\_1](https://doi.org/10.1007/0-306-47531-6_1)
- Walter, T., & Schuhfried, G. (2004). Computergestützte psychologische Diagnostik. In G. Mehta (Hrsg.), Springer eBook Collection Humanities, Social Science. *Die Praxis der Psychologie: Ein Karriereplaner* (S. 265-272). Wien: Springer. [https://doi.org/10.1007/978-3-7091-0571-9\\_21](https://doi.org/10.1007/978-3-7091-0571-9_21)
- Way, W. D. (2005). *Practical Questions in Introducing Computerized Adaptive Testing for K-12 Assessments: Research Report 05-03*. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=96c8892e542ac8dd46c532aa6400271e25b1c9d3> [08.01.2024]
- Weiss, D. J., & Kingsbury, G. G. (1984). Application of Computerized Adaptive Testing to Educational Problems. *Journal of Educational Measurement*, 21(4), 361-375. <http://www.jstor.org/stable/1434587>

Dr. Nikola Ebenbeck  
Ludwig-Maximilians-Universität München,  
Leopoldstraße 13, 80802 München  
[nikola.ebenbeck@lmu.de](mailto:nikola.ebenbeck@lmu.de)

Sven Anderson  
Technische Universität Dortmund,  
Emil-Figge-Straße 50, 44227 Dortmund

Prof. Dr. Jana Jungjohann  
Technische Universität Dortmund,  
Emil-Figge-Straße 50, 44227 Dortmund

Prof. Dr. Markus Gebhardt  
Ludwig-Maximilians-Universität München,  
Leopoldstraße 13, 80802 München